

(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開平11-259521

(43)公開日 平成11年(1999) 9月24日

(51)Int.Cl.⁶

識別記号

F I

G 0 6 F 17/30

G 0 6 F 15/401

3 2 0 A

17/27

15/38

D

15/40

3 7 0 A

審査請求 未請求 請求項の数21 O L (全 15 頁)

(21)出願番号 特願平10-63245

(22)出願日 平成10年(1998) 3月13日

(71)出願人 000005223

富士通株式会社

神奈川県川崎市中原区上小田中4丁目1番
1号

(72)発明者 西澤 信一郎

神奈川県川崎市中原区上小田中4丁目1番
1号 富士通株式会社内

(74)代理人 弁理士 伊東 忠彦

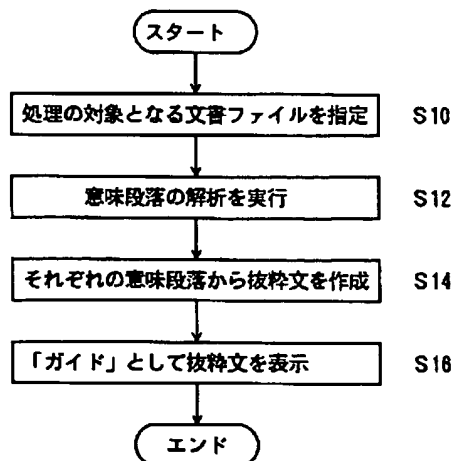
(54)【発明の名称】 文書理解支援装置、要約文生成方法、並びに文書理解支援プログラムを記録したコンピュータ読み取り可能な記録媒体

(57)【要約】

【課題】 システム規模が小さく処理時間が短くて済み、ガイド表示の抜粋文から元の文書の内容を参照することができ、文書を理解するための支援を効果的に行うことができる文書理解支援装置、要約文生成方法、並びに文書理解支援プログラムを記録したコンピュータ読み取り可能な記録媒体を提供することを目的とする。

【解決手段】 対象となる文書を、前記文書の内容を反映した段落構造である意味段落に分割する意味段落解析手段S12と、文書を分割した意味段落毎に意味段落に含まれる各文章の重要度を算出して重要度の高い文章を抜粋文として抽出する抜粋文抽出手段S14とを有する。このため、文書の構文解析を必要とせず、抜粋文を得るための統計量が少ないので、小さなシステム規模で、かつ、短い処理時間で文書理解を支援することができる。

本発明の文書理解支援システムの一実施例のフローチャート



【特許請求の範囲】

【請求項 1】 対象となる文書を、前記文書の内容を反映した段落構造である意味段落に分割する意味段落解析手段と、

前記文書を分割した意味段落毎に意味段落に含まれる各文章の重要度を算出して重要度の高い文章を抜粋文として抽出する抜粋文抽出手段とを有することを特徴とする文書理解支援装置。

【請求項 2】 請求項 1 記載の文書理解支援装置において、

前記抜粋文を意味段落単位で表示する抜粋文表示手段を有することを特徴とする文書理解支援装置。

【請求項 3】 請求項 2 記載の文書理解支援装置において、

前記抜粋文表示手段で表示された抜粋文の意味段落に対応する部分の文書を文書表示部に表示させ、前記文書表示部に表示された意味段落に対応する抜粋文を前記抜粋文表示手段に表示させる対応表示手段を有することを特徴とする文書理解支援装置。

【請求項 4】 請求項 1 または 2 記載の文書理解支援装置において、

前記意味段落解析手段は、

所定の規則に基づいて文章にポイントを与えるポイント手段と、

前記ポイント手段で与えられたポイントの合計値が所定値を超えた位置を前記意味段落の区切りとする意味段落決定手段とを有することを特徴とする文書理解支援装置。

【請求項 5】 請求項 4 記載の文書理解支援装置において、

前記ポイント手段は、

前記文書での出現頻度が高く集中して出現する度合いである重要度が高い重要語を抽出し、前記重要語が連続して出現する領域の先頭及び末尾の文章にポイントを与える第 1 のポイント手段を有することを特徴とする文書理解支援装置。

【請求項 6】 請求項 4 記載の文書理解支援装置において、

前記ポイント手段は、

複数ファイルで前記文書が構成されているとき、各ファイルの区切りの直後の文章にポイントを与える第 2 のポイント手段を有することを特徴とする文書理解支援装置。

【請求項 7】 請求項 4 記載の文書理解支援装置において、

前記ポイント手段は、

前記文書内の見出しの段落の直後の文章にポイントを与える第 3 のポイント手段を有することを特徴とする文書理解支援装置。

【請求項 8】 請求項 4 記載の文書理解支援装置におい

て、

前記ポイント手段は、

前記文書内の話題の転換を促す語で始まる文章にポイントを与える第 4 のポイント手段を有することを特徴とする文書理解支援装置。

【請求項 9】 請求項 4 記載の文書理解支援装置において、

前記ポイント手段は、

前記文書内の話題のつながりを促す語で始まる文章のポイントを減じる第 5 のポイント手段を有することを特徴とする文書理解支援装置。

【請求項 10】 請求項 4 乃至 9 のいずれか記載の文書理解支援装置において、

前記抜粋文抽出手段は、

前記意味段落の各文章に含まれる重要語についてその語の重要度を加算して文章の重要度を算出し、前記文章の重要度が大きいものから順に抜粋文として抽出することを特徴とする文書理解支援装置。

【請求項 11】 請求項 2 乃至 10 のいずれか記載の文書理解支援装置において、

前記抜粋文表示手段は、

任意の意味段落の抜粋文と共に、その意味段落と類似する意味段落の抜粋文を表示することを特徴とする文書理解支援装置。

【請求項 12】 コンピュータを、

対象となる文書を、前記文書の内容を反映した段落構造である意味段落に分割する意味段落解析手段と、

前記文書を分割した意味段落毎に意味段落に含まれる各文章の重要度を算出して重要度の高い文章を抜粋文として抽出する抜粋文抽出手段と、

前記抜粋文を意味段落単位で表示させる抜粋文表示手段と

して機能させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体。

【請求項 13】 請求項 12 記載のプログラムを記録したコンピュータ読み取り可能な記録媒体において、

前記抜粋文を意味段落単位で表示する抜粋文表示手段として機能させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体。

【請求項 14】 請求項 13 記載のプログラムを記録したコンピュータ読み取り可能な記録媒体において、

前記抜粋文表示手段で表示された抜粋文の意味段落に対応する部分の文書を文書表示部に表示させ、前記文書表示部に表示された意味段落に対応する抜粋文を前記抜粋文表示手段に表示させる対応表示手段として機能させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体。

【請求項 15】 文書から文章を抜粋して要約文を生成する要約文生成方法であって、

対象となる文書を、前記文書の内容を反映した段落構造

3

である意味段落に分割し、

前記文書を分割した意味段落毎に意味段落に含まれる各文章の重要度を算出して重要度の高い文章を抽出することを特徴とする要約文生成方法。

【請求項16】 請求項15記載の要約文生成方法であって、

所定の規則に基づいて文章にポイントを与え、

前記ポイント手段で与えられたポイントの合計値が所定値を超えた位置を前記意味段落の区切りとすることを特徴とする要約文生成方法。

【請求項17】 請求項16記載の要約文生成方法であって、

前記文書での出現頻度が高く集中して出現する度合いである重要度が高い重要語を抽出し、

前記重要語が連続して出現する領域の先頭及び末尾の文章にポイントを与えることを特徴とする要約文生成方法。

【請求項18】 請求項16乃至17のいずれか記載の要約文生成方法であって、

前記文書が複数ファイルで構成されているとき、各ファイルの区切りの直後の文章にポイントを与えることを特徴とする要約文生成方法。

【請求項19】 請求項16乃至18のいずれか記載の要約文生成方法であって、

前記文書内の見出しの段落の直後の文章にポイントを与えることを特徴とする要約文生成方法。

【請求項20】 請求項16乃至19のいずれか記載の要約文生成方法であって、

前記文書内の話題の転換を促す語で始まる文章にポイントを与えることを特徴とする要約文生成方法。

【請求項21】 請求項16乃至20のいずれか記載の要約文生成方法であって、

前記文書内の話題のつながりを促す語で始まる文章のポイントを減じることを特徴とする要約文生成方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は文書理解支援装置、要約文生成方法、並びに文書理解支援プログラムを記録したコンピュータ読み取り可能な記録媒体に係り、文書の内容を把握する手助けとなる要約文を生成し管理する文書理解支援装置、要約文生成方法、並びに文書理解支援プログラムを記録したコンピュータ読み取り可能な記録媒体に関する。

【0002】

【従来の技術】従来より、利用者が文書の内容を把握する手助けとなる情報を提供する機能として、その文書の内容を良く表している文章を抽出し、要約として表示したり保存するシステムが提案されている。例えば特開平3-191475号公報には、文書を見出しの付いた段落単位に分割し、段落の種類に対応した重要文抽出規則

(3)

特開平11-259521

4

群を段落毎に選択し、これと段落の種類に依存しない重要文抽出規則群の双方を参照して各段落における重要文を抽出し、各段落で抽出された重要文と各段落の見出しの対の集合を文書全体の重要部分と判定し要約として出力することが記載されている。

【0003】また、特開平2-254566号公報には、文書の構文解析を行って得られる形式的な段落構造を最小単位とし、新出語率と用語類似度を用いて形式的な段落構造をまとめてできる範囲を意味段落としてとらえ、更に、使用度数の他7種類の統計量を用いて重要語及び重要文を抽出し、これらから文書の抄録を得ることが記載されている。

【0004】また、特開昭63-298563号公報には、文章の形式、内容を参考にしてユーザが文章の要約範囲を指定し、各部分に対して要約処理方法を指定して、要約処理部で各分割部分の要約を行い、これを所定の順序でまとめ要約文章として出力することが記載されている。

【0005】

【発明が解決しようとする課題】従来の特開平3-191475号公報に記載のものでは、文書を見出しで区切られた段落構造に分けて各段構造の見出しに書かれている内容から種類を判定し重要文抽出の規則を選択しており、各段落の種類に依存した意味で重要文を抽出し段落の見出しと対応付けて表示するため、この表示内容は元の文書の流れに沿ったものとなる。このため、元の文書の流れとは違った視点から文書内容を確認することができないという問題があった。また、要約の文章から元の文書の内容を参照することはできず、使い勝手が悪いという問題があった。

【0006】また、特開平2-254566号公報に記載のものは、文書の構文解析を行って得られる形式的な段落構造を最小単位とするため、規則群や解析システムを準備しなければならず、システム全体の規模が大きくなり、処理に長時間を要する。また、意味段落に分割するために新出語率と用語類似度を算出し、重要語及び重要文を抽出するために使用度数の他7種類の統計量を算出しており、このために多大の計算量及び計算時間が必要となるという問題があった。

【0007】また、特開昭63-298563号公報に記載のものは、文章の要約範囲の指定及び各部分の要約処理方法の指定はユーザが行っており、全て自動で要約を作成することができないという問題があった。本発明は、上記の点に鑑みなされたもので、システム規模が小さく処理時間が短くて済み、ガイド表示の抜粋文から元の文書の内容を参照することができ、文書を理解するための支援を効果的に行うことができる文書理解支援装置、要約文生成方法、並びに文書理解支援プログラムを記録したコンピュータ読み取り可能な記録媒体を提供することを目的とする。

【0008】

【課題を解決するための手段】請求項1に記載の発明は、対象となる文書を、前記文書の内容を反映した段落構造である意味段落に分割する意味段落解析手段と、前記文書を分割した意味段落毎に意味段落に含まれる各文章の重要度を算出して重要度の高い文章を抜粋文として抽出する抜粋文抽出手段とを有する。

【0009】このように、文書その内容を反映した段落構造である意味段落に分割し、分割した意味段落毎に意味段落に含まれる各文章の重要度を算出して重要度の高い文章を抜粋文として抽出し、意味段落単位で表示するため、文書の構文解析を必要とせず、抜粋文を得るための統計量が少ないので、小さなシステム規模で、かつ、短い処理時間で文書理解を支援することができる。

【0010】請求項2に記載の発明は、請求項1記載の文書理解支援装置において、前記抜粋文を意味段落単位で表示する抜粋文表示手段を有する。このように、抜粋文を意味段落単位で表示することにより、文書理解を支援する抜粋文を表示できる。請求項3に記載の発明は、請求項2記載の文書理解支援装置において、前記抜粋文表示手段で表示された抜粋文の意味段落に対応する部分の文書を文書表示部に表示させ、前記文書表示部に表示された意味段落に対応する抜粋文を前記抜粋文表示手段に表示させる対応表示手段を有する。

【0011】このように、表示されている文書に対応する抜粋文を表示できると共に、表示されている抜粋文に対応する元の文書を表示でき、必要な情報を元の文書から早急に探し出すことができる。請求項4に記載の発明は、請求項1または2記載の文書理解支援装置において、前記意味段落解析手段は、所定の規則に基づいて文章にポイントを与えるポイント手段と、前記ポイント手段で与えられたポイントの合計値が所定値を超えた位置を前記意味段落の区切りとする意味段落決定手段とを有する。

【0012】このように、文章にポイントを与え、ポイントの合計値が所定値を超えた位置を意味段落の区切りとすることにより、話題の範囲毎の意味段落を効率的に見つけることができる。請求項5に記載の発明は、請求項4記載の文書理解支援装置において、前記ポイント手段は、前記文書での出現頻度が高く集中して出現する度合いである重要度が高い重要語を抽出し、前記重要語が連続して出現する領域の先頭及び末尾の文章にポイントを与える第1のポイント手段を有する。

【0013】このように、文書での出現頻度が高く集中して出現する度合いである重要度が高い重要語を抽出し、重要語が連続して出現する領域の先頭及び末尾の文章にポイントを与えることにより、話題の範囲毎の意味段落を効率的に見つけることができる。請求項6に記載の発明は、請求項4記載の文書理解支援装置において、前記ポイント手段は、複数ファイルで前記文書が構成さ

(4)

特開平11-259521

6

れているとき、各ファイルの区切りの直後の文章にポイントを与える第2のポイント手段を有する。

【0014】このように、ファイルの区切りの直後の文章にポイントを与えることにより、複数ファイルで構成された文書での話題の範囲毎の意味段落を確実に見つけることができる。請求項7に記載の発明は、請求項4記載の文書理解支援装置において、前記ポイント手段は、前記文書内の見出しの段落の直後の文章にポイントを与える第3のポイント手段を有する。

10 【0015】このように、文書内の見出しの段落の直後の文章にポイントを与えることにより、話題の範囲毎の意味段落を確実に簡単に見つけることができる。請求項8に記載の発明は、請求項4記載の文書理解支援装置において、前記ポイント手段は、前記文書内の話題の転換を促す語で始まる文章にポイントを与える第4のポイント手段を有する。

20 【0016】このように、話題の転換を促す語で始まる文章にポイントを与えることにより、話題の範囲毎の意味段落を確実に簡単に見つけることができる。請求項9に記載の発明は、請求項4記載の文書理解支援装置において、前記ポイント手段は、前記文書内の話題のつながりを促す語で始まる文章のポイントを減じる第5のポイント手段を有する。

30 【0017】このように、話題のつながりを促す語で始まる文章のポイントを減じることにより、意味段落ではない部分が誤って話題の範囲毎の意味段落となることを防止できる。請求項10に記載の発明は、請求項4乃至9のいずれか記載の文書理解支援装置において、前記抜粋文抽出手段は、前記意味段落の各文章に含まれる重要語についてその語の重要度を加算して文章の重要度を算出し、前記文章の重要度が大きいものから順に抜粋文として抽出する。

40 【0018】このように、意味段落の各文章に含まれる重要語についてその語の重要度を加算して文章の重要度を算出し、この重要度が大きいものから順に抜粋文として抽出するため、話題の範囲毎の意味段落を見つけるために使った語の重要度を再度利用して、意味段落の話題の内容を良く表す抜粋文を効率的に抽出できる。請求項11に記載の発明は、請求項2乃至10のいずれか記載の文書理解支援装置において、前記抜粋文表示手段は、任意の意味段落の抜粋文と共に、その意味段落と類似する意味段落の抜粋文を表示する。

50 【0019】このように、意味段落の抜粋文と共に、その意味段落と類似する意味段落の抜粋文を表示するため、類似する意味段落の抜粋文を用いて、文書内で類似した話題に関する部分をまとめて読むことができる。請求項12に記載の発明は、コンピュータを、対象となる文書を、前記文書の内容を反映した段落構造である意味段落に分割する意味段落解析手段と、前記文書を分割した意味段落毎に意味段落に含まれる各文章の重要度を算

出して重要度の高い文章を抜粋文として抽出する抜粋文抽出手段と、前記抜粋文を意味段落単位で表示させる抜粋文表示手段として機能させるためのプログラムを記録している。

【0020】この記録媒体を使用することにより、文書理解支援機能を持たせることができ、これにより請求項1記載の発明を実現できる。請求項13に記載の発明は、請求項12記載のプログラムを記録したコンピュータ読み取り可能な記録媒体において、前記抜粋文を意味段落単位で表示する抜粋文表示手段として機能させるためのプログラムを記録している。

【0021】この記録媒体を使用することにより、文書理解支援機能を持たせることができ、これにより請求項2記載の発明を実現できる。請求項14に記載の発明は、請求項13記載のプログラムを記録したコンピュータ読み取り可能な記録媒体において、前記抜粋文表示手段で表示された抜粋文の意味段落に対応する部分の文書を文書表示部に表示させ、前記文書表示部に表示された意味段落に対応する抜粋文を前記抜粋文表示手段に表示させる対応表示手段として機能させるためのプログラムを記録している。

【0022】この記録媒体を使用することにより、文書理解支援機能を持たせることができ、これにより請求項3記載の発明を実現できる。請求項15に記載の発明は、文書から文章を抜粋して要約文を生成する要約文生成方法であって、対象となる文書を、前記文書の内容を反映した段落構造である意味段落に分割し、前記文書を分割した意味段落毎に意味段落に含まれる各文章の重要度を算出して重要度の高い文章を抽出する。

【0023】このように、文書をその内容を反映した段落構造である意味段落に分割し、分割した意味段落毎に意味段落に含まれる各文章の重要度を算出して重要度の高い文章を抜粋文として抽出し、意味段落単位で表示するため、文書の構文解析を必要とせず、抜粋文を得るための統計量が少ないので、小さなシステム規模で、かつ、短い処理時間で文書理解を支援することができる。

【0024】請求項16に記載の発明は、請求項15記載の要約文生成方法であって、所定の規則に基づいて文章にポイントを与え、前記ポイント手段で与えられたポイントの合計値が所定値を超えた位置を前記意味段落の区切りとする。このように、文章にポイントを与え、ポイントの合計値が所定値を超えた位置を意味段落の区切りとすることにより、話題の範囲毎の意味段落を効率的に見つけることができる。

【0025】請求項17に記載の発明は、請求項16記載の要約文生成方法であって、前記文書での出現頻度が高く集中して出現する度合いである重要度が高い重要語を抽出し、前記重要語が連続して出現する領域の先頭及び末尾の文章にポイントを与える。

【0026】このように、文書での出現頻度が高く集中

して出現する度合いである重要度が高い重要語を抽出し、重要語が連続して出現する領域の先頭及び末尾の文章にポイントを与えることにより、話題の範囲毎の意味段落を効率的に見つけることができる。請求項18に記載の発明は、請求項16乃至17のいずれか記載の要約文生成方法であって、前記文書が複数ファイルで構成されているとき、各ファイルの区切りの直後の文章にポイントを与える。

【0027】このように、ファイルの区切りの直後の文章にポイントを与えることにより、複数ファイルで構成された文書での話題の範囲毎の意味段落を確実に見つけることができる。請求項19に記載の発明は、請求項16乃至18のいずれか記載の要約文生成方法であって、前記文書内の見出しの段落の直後の文章にポイントを与える。

【0028】このように、文書内の見出しの段落の直後の文章にポイントを与えることにより、話題の範囲毎の意味段落を確実に簡単に見つけることができる。請求項20に記載の発明は、請求項16乃至19のいずれか記載の要約文生成方法であって、前記文書内の話題の転換を促す語で始まる文章にポイントを与える。

【0029】このように、話題の転換を促す語で始まる文章にポイントを与えることにより、話題の範囲毎の意味段落を確実に簡単に見つけることができる。請求項21に記載の発明は、請求項16乃至20のいずれか記載の要約文生成方法であって、前記文書内の話題のつながりを促す語で始まる文章のポイントを減じる。

【0030】このように、話題のつながりを促す語で始まる文章のポイントを減じることにより、意味段落ではない部分が誤って話題の範囲毎の意味段落となることを防止できる。

【0031】

【発明の実施の形態】図1は本発明の文書理解支援装置の一実施例のブロック図を示す。同図中、中央処理装置(CPU)10には、バス15を介して入力装置20、記憶装置30、表示装置40、印刷装置50それぞれが接続されている。入力装置20としてはキーボード21、マウス22、スキャナ23等が設けられており、記憶装置30としてはRAM31、ROM32、ハードディスク装置33、フレキシブルディスク装置34等が設けられている。CPU10は記憶装置30に記憶されている後述の処理を行う各種処理プログラムを実行し、その結果を記憶装置30に記憶すると共に、表示装置40に表示し、印刷装置50で印刷して出力する。また、記憶装置30には各種処理プログラムの他に各種ライブラリも記憶されている。

【0032】図2は本発明の文書理解支援装置が実行する文書理解支援処理の一実施例のフローチャートを示す。同図中、ステップS10では入力装置20を用いて記憶装置30に格納されている文書ファイルから処理の

対象となる文書の文書ファイルを指定する。ここでは、1つの文書が複数の文書ファイルに分かれている場合には、該当する複数の文書ファイルを指定する。次に、ステップS12で指定した文書について話題範囲（意味段落）の解析を行う。次に、ステップS14でそれぞれの意味段落から抜粋文を作成し、ステップS16で作成した抜粋文をガイドとして表示する。

【0033】図3はステップS12で実行する意味段落解析ルーチンの一実施例のフローチャートを示す。同図中、ステップS20で文書を一定の文章数毎に区切り、複数の重要語判定領域に分割する。次に、ステップS22で各重要語判定領域毎に名詞の重要度を決定し、この各重要語判定領域毎の名詞の重要度を集めて文書全体の重要語を決定する。

【0034】ここでは、 $tf = (\text{その重要語判定領域でのその名詞の出現回数}) / (\text{文書全体でのその名詞の出現回数})$ 、 $idf = (\text{全体の重要語判定領域数}) / (\text{文書全体でその名詞が出現する重要語判定領域数})$ として、その名詞の重要度 $= tf \cdot idf$ とする。この計算式は文書での出現頻度が高いということと、ある領域に集中して出現するということが重要度を評価している。次に、各重要語判定領域の内部で重要度が所定の閾値を超えるものを各重要語判定領域の重要語と決定し、各重要語判定領域の重要語の集合を作成する。そして、各重要語判定領域の重要語の集合を文書全体でまとめることにより文書全体の重要語の集合を作成する。

【0035】次のステップS24で文書の1文章毎のループ処理で、文章毎に得点を与える。ここでは、ステップS26で、文書が複数の文書ファイルで構成されるときは各文書ファイルの区切りを意味段落の区切りとするため、その文章の直前で文書ファイルが始まっていると例えば2ポイント加点する。また、文書中の章、節等の見出しの段落が存在する位置を意味段落の区切りとするため、その文章の直前に章、節等の見出しの段落が存在すると例えば2ポイント加点する。また、「はじめに」、「さて」、「一方」、「ところで」等の話題の転換を促す語の始まる位置を意味段落の区切りとみなすため、話題の転換を促す語で始まる文章は例えば2ポイント加点する。また、「このように」、「そうすると」等の話題が続くことを示す語の始まる位置を意味段落の区切りではないとみなすため、話題が続くことを示す語で始まる文章は例えば2ポイント減点する。

【0036】次のステップS28は文章中の各名詞についてのループ処理で、文章毎に得点を与える。ここでは、ステップS30で文書全体の重要語を含んだ文章が連続する範囲の先頭位置を意味段落の区切りとするため、全ての名詞について、その名詞が直前の文章に現れてないかどうかを判別し、その名詞が直前の文章に現れてない場合にのみステップS32に進む。ステップS32ではその名詞が文書全体の重要語かつ重要語判定領域

の重要語であれば例えば1.5ポイント加点し、その名詞が文書全体の重要語だけであれば例えば1ポイント加点する。そして、ステップS34でステップS28のループ処理の終了が検出されると次のステップS36に進む。

【0037】ステップS36は文章中に現れなかった文書全体の重要語に属する各名詞についてのループ処理で、文章毎に得点を与える。ここでは、ステップS38で文書全体の重要語を含んだ文章が連続する範囲の末尾位置を意味段落の区切りとするため、その文書全体の重要語に属する名詞が直前の文章に現れていたかどうかを判別し、その文書全体の重要語に属する名詞が直前の文章に現れていた場合にのみステップS40に進む。ステップS40ではその文書全体の重要語に属する名詞が更にその重要語判定領域の重要語であれば例えば1.5ポイント加点し、その名詞が文書全体の重要語だけであれば例えば1ポイント加点する。

【0038】そして、ステップS42でステップS36のループ処理の終了が検出されると次のステップS44に進む。ここでステップS24のループ処理の終了が検出されるとステップS46に進む。ここでは、合計ポイントが所定値を超えた文章を意味段落の開始または末尾の区切りとみなし、各意味段落領域を決定して処理サイクルを終了する。なお、このステップS46で意味段落領域が決定されるのに応じて、重要語判定領域の重要語の集合から意味段落領域の重要語の集合が再編成される。

【0039】これによって、図4に示すように、文書が複数の文書ファイル60～62で構成されるときは各文書ファイル60～62の区切りが意味段落の境界となる。また、図5に示すように、文書中の章、節等の見出しの段落が存在する位置が意味段落の境界となる可能性が高い。また、図6に示すように、「はじめに」、「さて」、「一方」、「ところで」等の話題の転換を促す語の始まる位置が意味段落となり、「このように」、「そうすると」等の話題が続くことを示す語の始まる位置は意味段落の境界となる可能性が低くなる。

【0040】更に、図7に示すように、名詞Aが連続して出現する範囲65、67と名詞Aが出現しない範囲66とが隣接していると、少なくとも名詞Aは重要語判定領域の重要語となるため、上記の範囲65と範囲66との間、範囲65と範囲66との間が意味段落の境界となる可能性が高い。なお、図8に文章毎にポイントを加算した様子を示す。ここでは、丸印がその文章の合計ポイントを示し、白四角はその文章とが前の文章との間に意味段落の境界がないことを表し、黒四角はその文章とが前の文章との間に意味段落の境界があることを表している。

【0041】図9及び図11はステップS14で実行する抜粋文作成ルーチンの一実施例のフローチャートを示

す。図9のルーチンは意味段落領域毎に実行される。同図中、ステップS50で文書全体の重要語それぞれと意味段落領域の重要語それぞれの名詞の重要度を保存しておく。つまり、重要度が保存されている名詞は重要語である。次に、ステップS52で文書の1文章毎のループ処理を行い、次のステップS54で文章中の各名詞についてのループ処理を行って、文章毎に重要度を与える。

【0042】ここでは、ステップS56でその名詞が重要語であるか、つまり、重要度が保存されている否かを判別する。そして、その名詞が重要語である場合にのみステップS60に進む。ステップS60ではその名詞の重要度をその文章の重要度に加算する。そして、ステップS62でステップS54のループ処理の終了が検出されると次のステップS64に進む。ここでステップS52のループ処理の終了が検出されると処理サイクルを終了する。

【0043】図10では、テーブル70の第1欄に、文章単位でその中に含まれる名詞を示している。なお、重要語は大文字で、非重要語は小文字で示している。重要語についてはテーブル72の重要度を加算してテーブル70の第2欄に示す重要度の総和が求められる。3番目の文章では重要語の名詞D、B、D、Dがあるために、これらの重要度の総和は30となる。

【0044】図11は図9に続いて実行する抜粋文作成ルーチンの一実施例のフローチャートを示す。同図中、ステップS70で1つの意味段落領域から抜粋文を何文ずつ選択してガイド表示するかについて、利用者が指定しておく。次に、ステップS72で文書全体に対して意味段落一つずつのループ処理を行い、次のステップS74で意味段落に含まれる1文章毎のループ処理で、抜粋文の抽出を行う。

【0045】ここでは、ステップS76で指定された数だけの文章が既に抽出保存されているか否かを判別する。そして、指定された数だけの文章が既に抽出保存されている場合はステップS78に進み、指定された数だけの文章が抽出保存されていない場合はステップS80に進む。ステップS80では現在処理中の文章を抽出して抜粋文保存領域に保存する。この抜粋文保存領域は記憶装置30に意味段落毎に設けられており、各抜粋文保存領域には対応する意味段落とのリンク情報（例えば文書ファイル名及び文書ファイル内アドレス）も格納されている。

【0046】ステップS78では抜粋文保存領域に保存されている文章つまり抜粋文の中で重要度が最小の抜粋文と比べて現在処理中の文章の重要度が大きいのか否かを判別する。そして、現在処理中の文章の重要度が大きい場合にのみステップS82に進む。ステップS82では抜粋文保存領域に保存されている文章つまり抜粋文の中で重要度が最小の抜粋文を削除して、現在処理中の文章を抽出して抜粋文保存領域に保存する。ステップS84

でステップS74のループ処理の終了が検出されると次のステップS86に進む。ここでステップS72のループ処理の終了が検出されると処理サイクルを終了する。

【0047】図12はステップS16で表示されるガイド表示の一実施例を示す。ここで、本文表示ウインドウ80には文書の本文が表示されている。ガイド表示ウインドウ82にガイド表示が行われる。また、ガイド表示ウインドウ82と共に、追跡メニュー84が表示される。このガイド表示ウインドウ82は文書理解支援処理のステップS16で表示される場合は第1の意味段落に対応する内容が表示され、この後、本文表示ウインドウ80のガイドボタン81をクリックすると本文表示ウインドウ80に表示されている意味段落に対応する内容のガイド表示ウインドウ82が表示される。また、追跡メニュー84はガイド表示ウインドウ82の追跡ボタン83をクリックすることで表示/消去の切り替えが可能とされている。

【0048】図13はガイド表示ウインドウ82の拡大図を示す。同図中、表示部85に意味段落単位で指定された意味段落（指定意味段落）の抜粋文と、これに類似する意味段落の抜粋文の一部が表示される。ここで、類似する意味段落とは、指定意味段落の重要語の集合にある重要語と同一の重要語を所定数以上含む意味段落である。この類似する意味段落が複数ある場合には同一の重要語を最も多く含む意味段落が選択されて表示部85に表示される。表示部85に表示された抜粋文はボタン86a、86bをクリックすることにより上下にスクロールし、また、ボタン87a、87bをクリックすることにより左右にスクロールする。

【0049】また、ガイド表示ウインドウ82には意味段落移動用のボタン88a、88bと元文書参照ボタン89が設けられている。意味段落移動用のボタン88aをクリックすることにより、図14(B)に示す現在表示中の指定意味段落に対するガイド表示から図14

(A)に示す直前の意味段落が指定意味段落としてガイド表示ウインドウ82に表示される。また、意味段落移動用のボタン88bをクリックすることにより、図14(B)に示す現在表示中の指定意味段落に対するガイド表示から図14(C)に示す直後の意味段落が指定意味段落としてガイド表示ウインドウ82に表示される。

【0050】元文書参照ボタン89をクリックすると、図15に示すように、ガイド表示ウインドウ82に現在表示されている指定意味段落に対応する部分の文書（元文書）が本文表示ウインドウ80に表示される。これによって、ガイド表示から元文書を参照することができ、追跡メニュー84にはボタン90a、90bが設けられている。図16に示すように、本文表示ウインドウ80に表示されている部分に対応する指定意味段落がガイド表示ウインドウ82に表示されているとき、ボタン90aをクリックすると、このガイド表示ウインドウ82

の指定意味段落に類似する意味段落のうち上記指定意味段落より前に存在する類似意味段落の部分を本文表示ウインドウ 80 に表示する。また、ボタン 90 b をクリックすると、このガイド表示ウインドウ 82 の指定意味段落に類似する意味段落のうち上記指定意味段落より後に存在する類似意味段落の部分を本文表示ウインドウ 80 に表示する。

【0051】図 17 に上記のガイド表示を行うときの操作の意志決定のフローチャートを示す。ここでは X 番目の意味段落を指定意味段落とするガイド表示がガイド表示ウインドウ 82 に表示されているものとする。同図中、ステップ S100 で X-1 番目の意味段落についてガイド表示したいか否かを判定し、そうであればステップ S102 でボタン 88 a をクリックする。そうでなければステップ S104 で X+1 番目の意味段落についてガイド表示したいか否かを判定し、そうであればステップ S106 でボタン 88 b をクリックする。

【0052】そうでなければステップ S108 で X 番目の意味段落に対応する部分の文書（元文書）を参照したいか否かを判定し、そうであればステップ S110 で元文書参照ボタン 89 をクリックする。そうでなければステップ S112 で X 番目の意味段落と内容が類似する意味段落に対応する部分の文書（元文書）を参照したいか否かを判定し、そうであればステップ S114 でボタン 86 a, 86 b, 87 a, 87 b をクリックしてスクロールを行い、所望の抜粋文を表示させて選択し、参照ボタン 89 をクリックする。

【0053】そうでなければステップ S116 で X 番目の意味段落と内容が類似する意味段落のうち X 番目より前の意味段落に対応する部分の文書（元文書）を参照したいか否かを判定し、そうであればステップ S118 で追跡メニュー 84 を表示し、ボタン 90 a をクリックする。そうでなければステップ S120 で X 番目の意味段落と内容が類似する意味段落のうち X 番目より後の意味段落に対応する部分の文書（元文書）を参照したいか否かを判定し、そうであればステップ S122 で追跡メニュー 84 を表示し、ボタン 90 b をクリックする。

【0054】このように、文書をその内容を反映した段落構造である意味段落に分割し、分割した意味段落毎に意味段落に含まれる各文章の重要度を算出して重要度の高い文章を抜粋文として保存し、意味段落単位で表示するため、文書の構文解析を必要とせず、抜粋文を得るための統計量が少ないので、小さなシステム規模で、かつ、短い処理時間で文書理解を支援する抜粋文を表示できる。また、表示されている文書に対応する抜粋文を表示できると共に、表示されている抜粋文に対応する元の文書を表示でき、必要な情報を元の文書から早急に探し出すことができる。

【0055】また、文書での出現頻度が高く集中して出現する度合いである重要度が高い重要語を抽出し、重要

語が連続して出現する領域の先頭及び末尾の文章にポイントを与え、ポイントの合計値が所定値を超えた位置を意味段落の区切りとすることにより、話題の範囲毎の意味段落を効率的に見つけることができ、ファイルの区切りの直後の文章にポイントを与えることにより、複数ファイルで構成された文書での話題の範囲毎の意味段落を確実に見つけることができ、文書内の見出しの段落の直後の文章にポイントを与えることにより、話題の範囲毎の意味段落を確実に簡単に見つけることができ、話題の転換を促す語で始まる文章にポイントを与えることにより、話題の範囲毎の意味段落を確実に簡単に見つけることができる。

【0056】また、話題のつながりを促す語で始まる文章のポイントを減じることにより、意味段落ではない部分が誤って話題の範囲毎の意味段落となることを防止でき、意味段落の各文章に含まれる重要語についてその語の重要度を加算して文章の重要度を算出し、この重要度が大きいものから順に抜粋文として保存するため、話題の範囲毎の意味段落を見つけるために使った語の重要度を再度利用して、意味段落の話題の内容を良く表す抜粋文を効率的に保存でき、意味段落の抜粋文と共に、その意味段落と類似する意味段落の抜粋文を表示するため、類似する意味段落の抜粋文を用いて、文書内で類似した話題に関する部分をまとめて読むことができる。

【0057】なお、ステップ S12 が意味段落解析手段に対応し、ステップ S14 が抜粋文抽出手段に対応し、ステップ S16 が抜粋文表示手段に対応する。また、本文表示ウインドウ 80, ガイド表示ウインドウ 82, ガイドボタン 81, 追跡ボタン 83 が対応表示手段に対応する。また、ステップ S26 が第 1～第 5 のポイント手段に対応する。

【0058】

【発明の効果】上述の如く、請求項 1 に記載の発明は、対象となる文書を、前記文書の内容を反映した段落構造である意味段落に分割する意味段落解析手段と、前記文書を分割した意味段落毎に意味段落に含まれる各文章の重要度を算出して重要度の高い文章を抜粋文として抽出する抜粋文抽出手段とを有する。

【0059】このように、文書をその内容を反映した段落構造である意味段落に分割し、分割した意味段落毎に意味段落に含まれる各文章の重要度を算出して重要度の高い文章を抜粋文として抽出し、意味段落単位で表示するため、文書の構文解析を必要とせず、抜粋文を得るための統計量が少ないので、小さなシステム規模で、かつ、短い処理時間で文書理解を支援することができる。

【0060】また、請求項 2 に記載の発明は、抜粋文を意味段落単位で表示する抜粋文表示手段を有する。このように、抜粋文を意味段落単位で表示することにより、文書理解を支援する抜粋文を表示できる。また、請求項 3 に記載の発明は、抜粋文表示手段で表示された抜粋文

の意味段落に対応する部分の文書を文書表示部に表示させ、前記文書表示部に表示された意味段落に対応する抜粋文を前記抜粋文表示手段に表示させる対応表示手段を有する。

【0061】このように、表示されている文書に対応する抜粋文を表示できると共に、表示されている抜粋文に対応する元の文書を表示でき、必要な情報を元の文書から早急に探し出すことができる。また、請求項4に記載の発明では、意味段落解析手段は、所定の規則に基づいて文章にポイントを与えるポイント手段と、前記ポイント手段で与えられたポイントの合計値が所定値を超えた位置を前記意味段落の区切りとする意味段落決定手段とを有する。

【0062】このように、文章にポイントを与え、ポイントの合計値が所定値を超えた位置を意味段落の区切りとすることにより、話題の範囲毎の意味段落を効率的に見つけることができる。また、請求項5に記載の発明では、ポイント手段は、前記文書での出現頻度が高く集中して出現する度合いである重要度が高い重要語を抽出し、前記重要語が連続して出現する領域の先頭及び末尾の文章にポイントを与える第1のポイント手段を有する。

【0063】このように、文書での出現頻度が高く集中して出現する度合いである重要度が高い重要語を抽出し、重要語が連続して出現する領域の先頭及び末尾の文章にポイントを与えることにより、話題の範囲毎の意味段落を効率的に見つけることができる。また、請求項6に記載の発明では、ポイント手段は、複数ファイルで前記文書が構成されているとき、各ファイルの区切りの直後の文章にポイントを与える第2のポイント手段を有する。

【0064】このように、ファイルの区切りの直後の文章にポイントを与えることにより、複数ファイルで構成された文書での話題の範囲毎の意味段落を確実に見つけることができる。また、請求項7に記載の発明では、ポイント手段は、前記文書内の見出しの段落の直後の文章にポイントを与える第3のポイント手段を有する。

【0065】このように、文書内の見出しの段落の直後の文章にポイントを与えることにより、話題の範囲毎の意味段落を確実に簡単に見つけることができる。また、請求項8に記載の発明では、ポイント手段は、前記文書内の話題の転換を促す語で始まる文章にポイントを与える第4のポイント手段を有する。

【0066】このように、話題の転換を促す語で始まる文章にポイントを与えることにより、話題の範囲毎の意味段落を確実に簡単に見つけることができる。また、請求項9に記載の発明では、ポイント手段は、前記文書内の話題のつながりを促す語で始まる文章のポイントを減じる第5のポイント手段を有する。

【0067】このように、話題のつながりを促す語で始

まる文章のポイントを減じることにより、意味段落ではない部分が誤って話題の範囲毎の意味段落となることを防止できる。また、請求項10に記載の発明では、抜粋文抽出手段は、前記意味段落の各文章に含まれる重要語についてその語の重要度を加算して文章の重要度を算出し、前記文章の重要度が大きいものから順に抜粋文として抽出する。

【0068】このように、意味段落の各文章に含まれる重要語についてその語の重要度を加算して文章の重要度を算出し、この重要度が大きいものから順に抜粋文として抽出するため、話題の範囲毎の意味段落を見つけるために使った語の重要度を再度利用して、意味段落の話題の内容を良く表す抜粋文を効率的に抽出できる。また、請求項11に記載の発明では、抜粋文表示手段は、任意の意味段落の抜粋文と共に、その意味段落と類似する意味段落の抜粋文を表示する。

【0069】このように、意味段落の抜粋文と共に、その意味段落と類似する意味段落の抜粋文を表示するため、類似する意味段落の抜粋文を用いて、文書内で類似した話題に関する部分をまとめて読むことができる。また、請求項12に記載の発明は、コンピュータを、対象となる文書を、前記文書の内容を反映した段落構造である意味段落に分割する意味段落解析手段と、前記文書を分割した意味段落毎に意味段落に含まれる各文章の重要度を算出して重要度の高い文章を抜粋文として抽出する抜粋文抽出手段と、前記抜粋文を意味段落単位で表示させる抜粋文表示手段として機能させるためのプログラムを記録している。

【0070】この記録媒体を使用することにより、文書理解支援機能を持たせることができ、これにより請求項1記載の発明を実現できる。また、請求項13に記載の発明は、請求項12記載のプログラムを記録したコンピュータ読み取り可能な記録媒体において、前記抜粋文を意味段落単位で表示する抜粋文表示手段として機能させるためのプログラムを記録している。

【0071】この記録媒体を使用することにより、文書理解支援機能を持たせることができ、これにより請求項2記載の発明を実現できる。また、請求項14に記載の発明は、請求項13記載のプログラムを記録したコンピュータ読み取り可能な記録媒体において、前記抜粋文表示手段で表示された抜粋文の意味段落に対応する部分の文書を文書表示部に表示させ、前記文書表示部に表示された意味段落に対応する抜粋文を前記抜粋文表示手段に表示させる対応表示手段として機能させるためのプログラムを記録している。

【0072】この記録媒体を使用することにより、文書理解支援機能を持たせることができ、これにより請求項3記載の発明を実現できる。また、請求項15に記載の発明は、文書から文章を抜粋して要約文を生成する要約文生成方法であって、対象となる文書を、前記文書の内

容を反映した段落構造である意味段落に分割し、前記文書を分割した意味段落毎に意味段落に含まれる各文章の重要度を算出して重要度の高い文章を抽出する。

【0073】このように、文書をその内容を反映した段落構造である意味段落に分割し、分割した意味段落毎に意味段落に含まれる各文章の重要度を算出して重要度の高い文章を抜粋文として抽出し、意味段落単位で表示するため、文書の構文解析を必要とせず、抜粋文を得るための統計量が少ないので、小さなシステム規模で、かつ、短い処理時間で文書理解を支援することができる。

【0074】また、請求項16に記載の発明は、所定の規則に基づいて文章にポイントを与え、前記ポイント手段で与えられたポイントの合計値が所定値を超えた位置を前記意味段落の区切りとする。このように、文章にポイントを与え、ポイントの合計値が所定値を超えた位置を意味段落の区切りとすることにより、話題の範囲毎の意味段落を効率的に見つけることができる。

【0075】また、請求項17に記載の発明は、文書での出現頻度が高く集中して出現する度合いである重要度が高い重要語を抽出し、前記重要語が連続して出現する領域の先頭及び末尾の文章にポイントを与える。このように、文書での出現頻度が高く集中して出現する度合いである重要度が高い重要語を抽出し、重要語が連続して出現する領域の先頭及び末尾の文章にポイントを与えることにより、話題の範囲毎の意味段落を効率的に見つけることができる。

【0076】また、請求項18に記載の発明は、文書が複数ファイルで構成されているとき、各ファイルの区切りの直後の文章にポイントを与える。このように、ファイルの区切りの直後の文章にポイントを与えることにより、複数ファイルで構成された文書での話題の範囲毎の意味段落を確実に見つけることができる。

【0077】また、請求項19に記載の発明は、文書内の見出しの段落の直後の文章にポイントを与える。このように、文書内の見出しの段落の直後の文章にポイントを与えることにより、話題の範囲毎の意味段落を確実に簡単に見つけることができる。また、請求項20に記載の発明は、文書内の話題の転換を促す語で始まる文章にポイントを与える。

【0078】このように、話題の転換を促す語で始まる文章にポイントを与えることにより、話題の範囲毎の意味段落を確実に簡単に見つけることができる。また、請求項21に記載の発明は、文書内の話題のつながりを促す語で始まる文章のポイントを減じる。このように、話題のつながりを促す語で始まる文章のポイントを減じることにより、意味段落ではない部分が誤って話題の範囲毎の意味段落となることを防止できる。

【図面の簡単な説明】

【図1】本発明の文書理解支援装置の一実施例のブロック図である。

【図2】本発明の文書理解支援装置が実行する文書理解支援処理の一実施例のフローチャートである。

【図3】意味段落解析ルーチンの一実施例のフローチャートである。

【図4】意味段落の境界を説明するための図である。

【図5】意味段落の境界を説明するための図である。

【図6】意味段落の境界を説明するための図である。

【図7】意味段落の境界を説明するための図である。

【図8】文章毎にポイントを加算蓄積した様子を示す図である。

【図9】抜粋文作成ルーチンの一実施例のフローチャートである。

【図10】抜粋文作成ルーチンを説明するための図である。

【図11】抜粋文作成ルーチンの一実施例のフローチャートである。

【図12】ガイド表示の一実施例を示す図である。

【図13】ガイド表示ウインドウ82の拡大図である。

【図14】意味段落移動用のボタン88a、88bをクリックしたときのガイド表示を説明するための図である。

【図15】元文書参照ボタン89をクリックしたときのガイド表示を説明するための図である。

【図16】追跡メニュー84を説明するための図である。

【図17】ガイド表示を行うときの操作の意志決定のフローチャートである。

【符号の説明】

10 中央処理装置 (CPU)

20 入力装置

21 キーボード

22 マウス

23 スキャナ

30 記憶装置

31 RAM

32 ROM

33 ハードディスク装置

34 フレキシブルディスク装置

40 表示装置

50 印刷装置

80 本文表示ウインドウ

81 ガイドボタン

82 ガイド表示ウインドウ

83 追跡ボタン

84 追跡メニュー

85 表示部

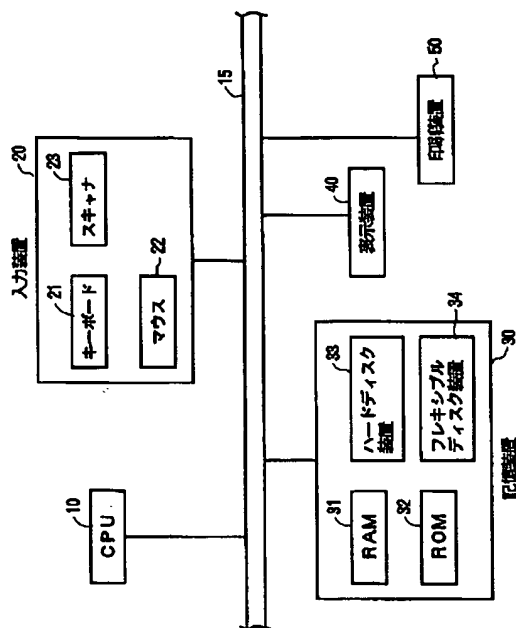
86a、86b、87a、87b、90a、90b ボタン

88a、88b 意味段落移動用のボタン

89 元文書参照ボタン

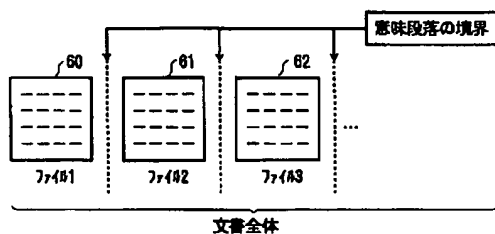
【図 1】

本発明の文書理解支援システムの一実施例のブロック図



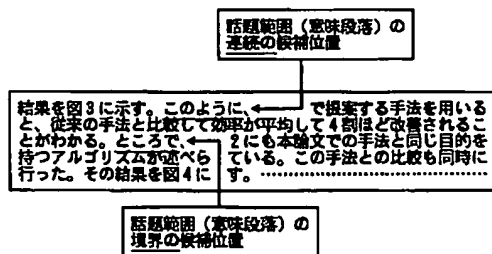
【図 4】

意味段落の境界を説明するための図



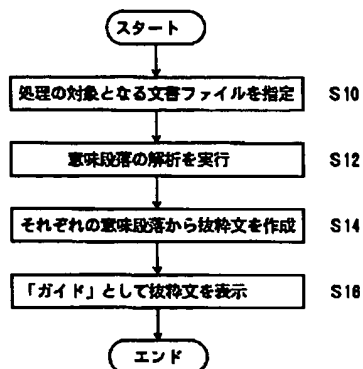
【図 6】

意味段落の境界を説明するための図



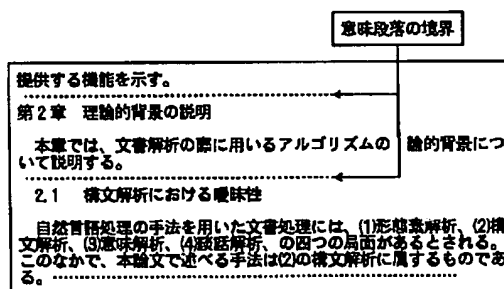
【図 2】

本発明の文書理解支援システムの一実施例のフローチャート



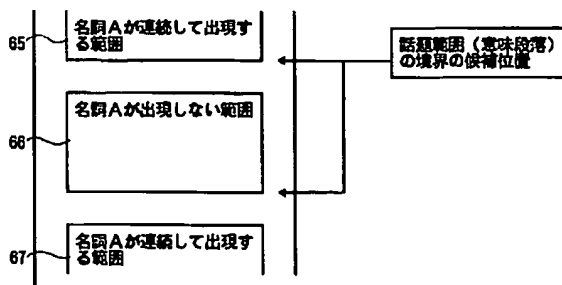
【図 5】

意味段落の境界を説明するための図



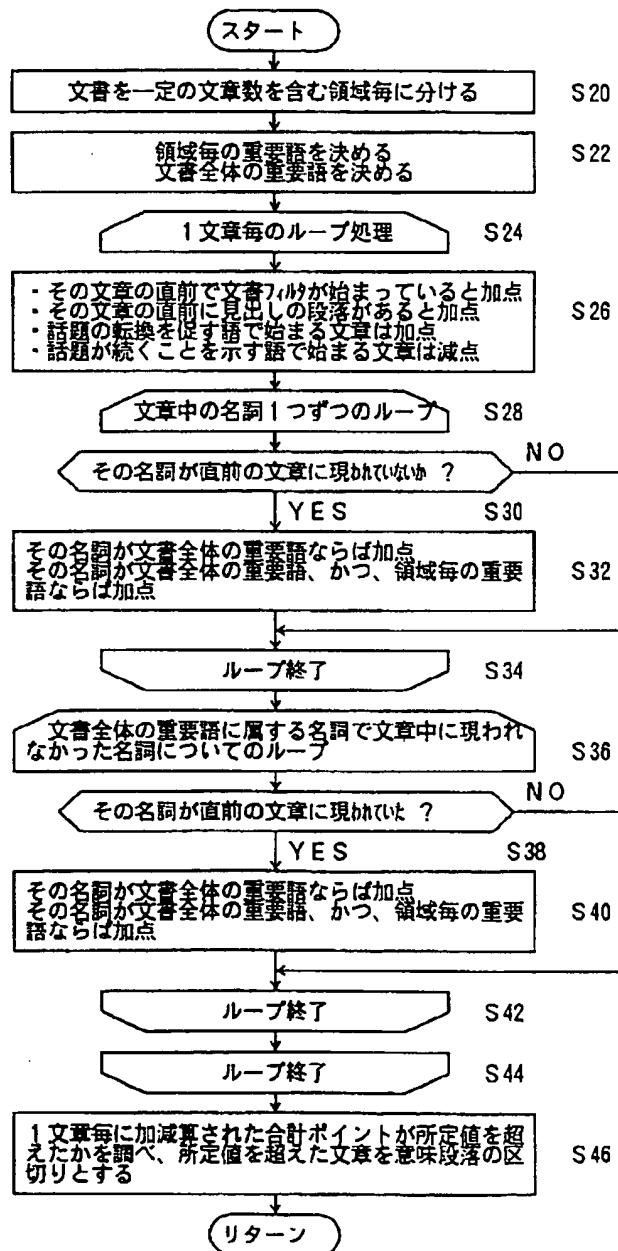
【図 7】

意味段落の境界を説明するための図



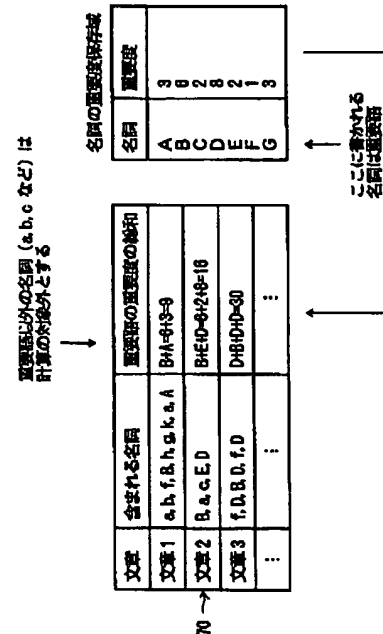
【図 3】

意味段落解析ルーチンの一実施例のフローチャート



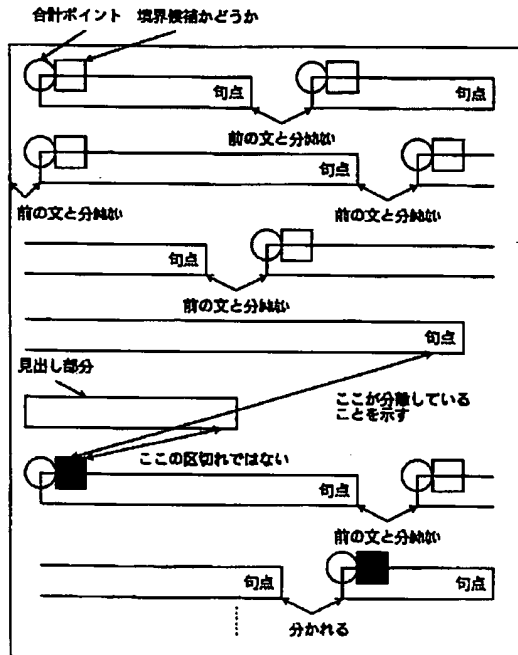
【図 10】

抜粋文作成ルーチンを説明するための図



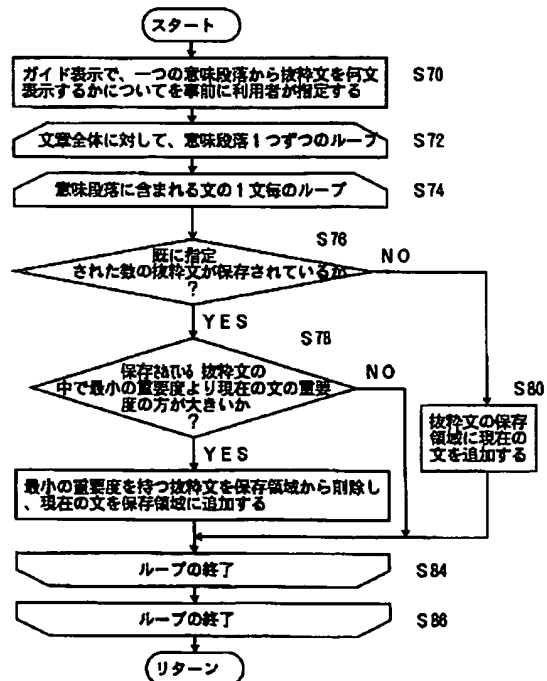
【図8】

文章毎にポイントを加算された様子を示す図



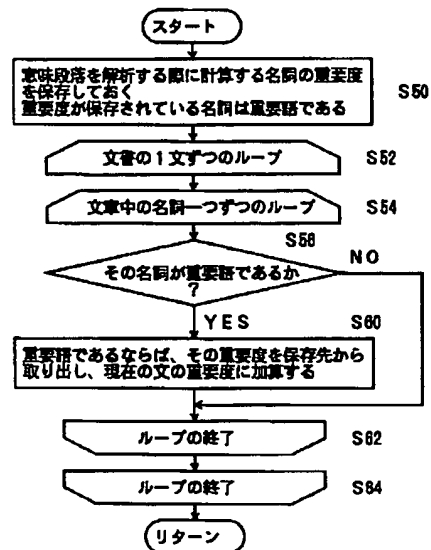
【図11】

抜粋作成ルーチンの一実施例のフローチャート



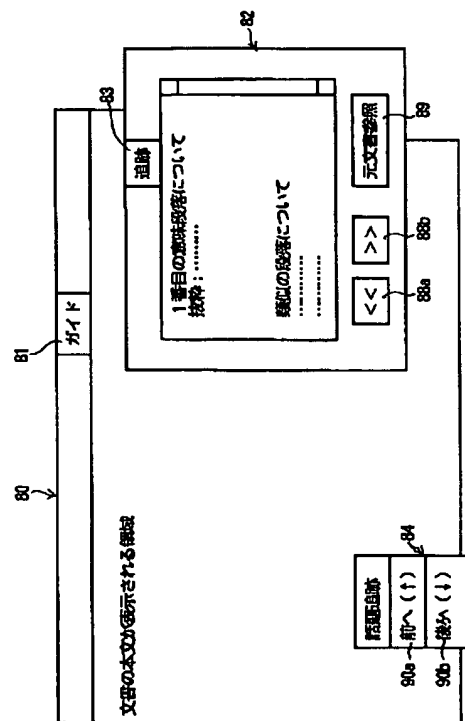
【図9】

抜粋作成ルーチンの一実施例のフローチャート



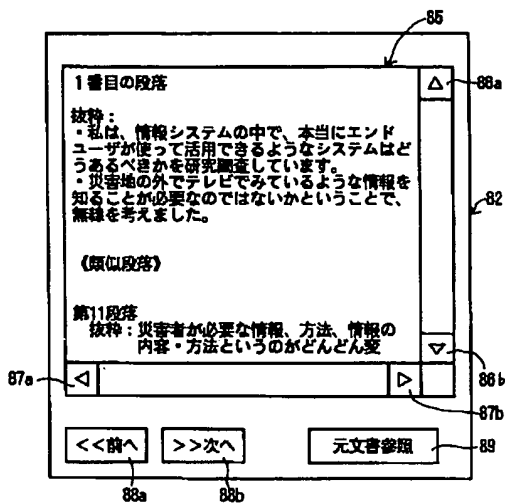
【図12】

ガイド表示ウィンドウ82の拡大図



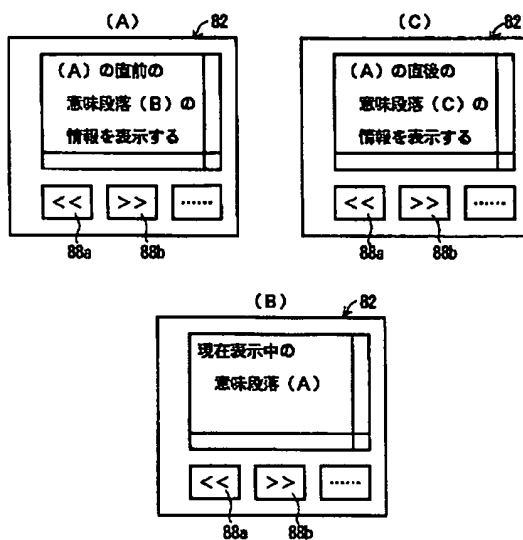
【図 13】

ガイド表示ウィンドウ 82 の拡大図



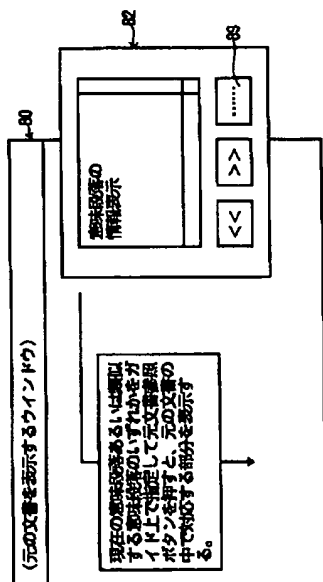
【図 14】

意味段落移動用のボタン 88a、88b をクリックしたときのガイド表示を説明するための図



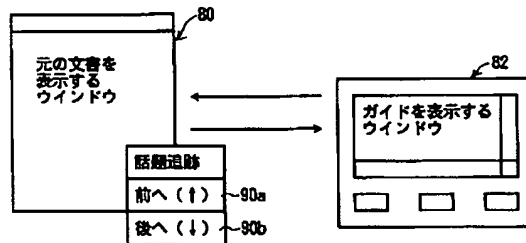
【図 15】

元文書参照ボタン 89 をクリックしたときのガイド表示を説明するための図



【図 16】

追跡メニュー 84 を説明するための図



【図17】

ガイド表示を行うときの操作の意志決定のフローチャート

